

El método de las palabras asociadas (I): La estructura de las redes científicas

ROSARIO RUIZ-BAÑOS
RAFAEL BAILÓN-MORENO
Universidad de Granada

El método de las palabras asociadas es una herramienta cuantitativa relacional de segunda generación que es capaz de poner de manifiesto las redes científicas a través de la co-ocurrencia de palabras en los documentos científicos. La red o mapa generado se divide en temas de investigación, definidos por sus descriptores y los parámetros densidad y centralidad. Igualmente permite disponer estos temas en un diagrama estratégico que nos representa tanto el desarrollo interno de los temas como su capacidad de relación con los demás dentro de la red científica.

PALABRAS CLAVE: Redes científicas. Método de las palabras asociadas. Leximappe. Centralidad. Densidad. Diagrama estratégico.

THE ASSOCIATED WORDS APPROACH: THE STRUCTURE OF SCIENTIFIC NETWORKS

The associated words approach is a second-generation relational scientometric tool capable of showing the existence of scientific networks through the co-occurrence of words in scientific documents. The generated network or map is divided into research topics, defined by their descriptors and the parameters of density and centrality. It also makes it possible to display these topics according to a strategic diagram representing both their internal development and their ability to relate with each other within the scientific network.

KEY WORDS: Scientific networks. Associated words approach. Leximappe. Centrality. Density. Strategic diagram.

1. INTRODUCCIÓN

La cuantimetría, generalización de la bibliometría a la globalidad de la ciencia, puede dividirse en dos categorías según se utilicen indicadores de actividad o indicadores de relación. Los indicadores de actividad nos informan sobre el volumen y el impacto de la actividad investigadora de las distintas áreas disciplinares, temáticas e incluso institucionales, tales como número de artículos científicos, productividad de los autores, factor de impacto de las revistas, y un larguísimo etcétera. En cambio, los indicadores relacionales pretenden delimitar las fronteras continuamente cambiantes de las disciplinas, conocer cómo están construidas, su evolución y su solapamiento con otras áreas científicas.

Son dos los análisis que nos conducen a indicadores relacionales: el análisis de co-citas o de primera generación, y el análisis de las palabras asociadas o de segunda generación. Tanto co-citas como palabras asociadas parten de la idea de que la ciencia está constituida por una red que entrelaza actores. Un actor puede ser un artículo científico, un investigador, una institución, un tema de investigación, un laboratorio e incluso un simple aparato de medida. Un actor es, en definitiva, cualquier entidad que genera nuevas redes ya que siempre presenta una capacidad mayor o menor de enlazamiento con otros actores.

El análisis de co-citas determina la estructura de esta red a través de la citación simultánea de autores, revistas o instituciones, que evidentemente, por esta circunstancia, deberán estar relacionados entre ellos. El inconveniente del análisis de co-citas es que no estudia de forma directa el conocimiento científico, los temas de investigación y además sólo es aplicable si se usan documentos recogidos por las bases de datos del ISI.

El método de las palabras asociadas es una herramienta cuantitativa desarrollada inicialmente en el Centre de Sociologie de l'Innovation (CSI) de l'Ecole Nationale Supérieure de Mines de Paris y en el Institut de l'Information Scientifique et Technique (antiguo CDST) del CNRS. Este método visualiza la estructura de las redes cuantitativas, de acuerdo con la teoría actor-red, y calcula una serie de parámetros que nos permiten estudiar el comportamiento de cada uno de los actores puestos en juego, tanto en su aspecto puramente estructural como en su aspecto evolutivo o dinámico.

Las bases teórico-metodológicas han sido establecidas por J. P. Courtial [1.986], que analiza las ventajas del método frente a un análisis escalar multidimensional, MDS, la estabilidad de los esqueletos de las redes en función del umbral y de la indización y otras cuestiones relativas a la pertinencia del análisis de asociaciones. Para la puesta en marcha de este método se ha desarrollado un conjunto de programas informáticos denominado LEXIMAPPE. En 1988, y en colaboración con la Universidad de Keele (Reino Unido), los programas originarios se modifican y mejoran para poder ser implementados en microordenadores tipo PC [Whittaker, 1988]. Ampliaciones sucesivas han sido desarrolladas en el CSI entre los años 1991 y 1992 [CSI, 1992] y por Bailón-Moreno para Ruiz-Baños [Ruiz-Baños, 1997].

Leximappe se aplica a todo tipo de documentos indizados mediante palabras clave y en especial a los artículos científicos y técnicos, patentes, etc. De forma más general es aplicable a cualquier documento textual, siempre y cuando se haga una indización semi-automática previa mediante un programa adecuado, tal como LEXINET, desarrollado igualmente por el CNRS [Chartron, 1989]. Por tanto, la gran ventaja que aporta el método de las palabras asociadas frente al análisis de co-citas radica en que el primero puede tomar la información de cualquier base de datos del mundo (por ejemplo PASCAL, FRANCIS, SCI, SSCI, AHCI, MEDLINE, BIOSIS, CHEMICAL ABSTRACTS, ICYT, ISOC, SIGLE, y un larguísimo etcétera),

mientras que el segundo está limitado prácticamente a la utilización de las bases SCI, SSCI, AHCI y ninguna más. La cobertura y posibilidades del análisis de palabras asociadas es prácticamente universal.

2. OBJETO

El método de las palabras asociadas, si bien ya es un instrumento habitual en países como Francia, Bélgica, Holanda e Inglaterra, es un gran desconocido dentro de nuestro país. Con la colaboración del Centre de Sociologie de l'Innovation (CSI) y a través de Jean-Pierre Courtial y de Anne Sigogneau, Rosario Ruiz Baños ha utilizado por primera vez en España este método aplicándolo a la Arqueología [Ruiz-Baños, 1997]. El alcance y originalidad de los resultados obtenidos han demostrado la bondad del análisis de palabras asociadas y siguiendo con la línea ya iniciada en otro trabajo anterior de divulgar las técnicas bibliométricas entre la comunidad de bibliotecarios y documentalistas [Ruiz-Baños et al., 1997], han movido a los autores a publicar los fundamentos de esta metodología, muy dispersos en la bibliografía de lengua no castellana, a través del presente artículo. En esta primera parte se mostrará la forma en que el método de las palabras asociadas es capaz de representar la estructura de cualquier red científica.

3. FUNDAMENTOS MATEMÁTICOS DEL MÉTODO DE LAS PALABRAS ASOCIADAS

3.1. Matriz de ocurrencias. Matriz de asociaciones

El método de las palabras asociadas considera que el contenido de un documento viene definido por sus descriptores o palabras clave. Se parte, por tanto, de una matriz de datos «documentos x palabras clave», denominada matriz de ocurrencias, que representaría el contenido conceptual del campo científico en estudio [Courtial, 1990].

La lista de palabras clave puede ser muy extensa, del orden de varios miles, por lo que las dimensiones de esta matriz de ocurrencias son extraordinarias. Si partimos de, por ejemplo, una base de datos con 10.000 documentos y manejamos un vocabulario de 5.000 palabras, el número de elementos que contendrá será de 50 millones (10.000 x 5.000). La ec. (1) representa una matriz de ocurrencias constituida por 3000 documentos y 1000 descriptores. El número de celdillas es de 3000 x 1000, que se completará con ceros y unos. Cuando un documento «i» contiene la palabra clave «j» en la celdilla «i,j» colocamos un 1 y en caso contrario se coloca un 0. Debido a que los documentos se indizan con muy pocas palabras clave, normalmente nunca más de 10 ó 15, la matriz de ocurrencias apenas si tiene celdillas ocupadas con unos. El número de veces que una palabra clave «j» aparece u ocurre se denota por C_j .

	<i>Pal. 1</i>	<i>Pal. 2</i>	... <i>Pal. i</i>	... <i>Pal. j</i>	... <i>Pal. 1000</i>
<i>Doc. 1</i>	0	0	... 0	... 1	... 0
<i>Doc. 2</i>	0	1	... 0	... 0	... 0
...
<i>Doc. i</i>	0	0	... 1	... 1	... 0
...
<i>Doc. j</i>	0	1	... 0	... 0	... 0
...
<i>Doc. 3000</i>	0	0	... 0	... 1	... 0
	1200	1350	... 58	... 720	... 250
	C_1	C_2	... C_i	... C_j	... C_{1000}

(1)

Según la ley de Zipf, la frecuencia de aparición de palabras en un texto es muy baja en la mayoría de los casos, por lo que la mayor parte de los descriptores serán poco abundantes y pueden ser despreciados. Por esta razón, LEXIMAPPE sólo reconoce como máximo hasta las primeras 1.500 palabras más frecuentes, siendo lo más usual utilizar unas 700.

Decimos que dos palabras co-ocurren cuando aparecen simultáneamente en el mismo documento. Dos palabras estarán más ligadas o asociadas entre sí cuanto mayor sea la co-ocurrencia entre ellas. Por tanto, la medida del enlace entre dos palabras de una red será proporcional a la co-ocurrencia de esas dos palabras en el conjunto de documentos que se tome como muestra.

La matriz de asociaciones, de co-ocurrencias o de «palabras clave x palabras clave» es una matriz de adyacencia cuadrada simétrica donde cada elemento representa la asociación entre los descriptores. En la celdilla C_{ij} colocamos el número de documentos en los que la palabra «i» y la palabra «j» aparecen simultáneamente.

	<i>Pal. 1</i>	<i>Pal. 2</i>	... <i>Pal. i</i>	... <i>Pal. j</i>	... <i>Pal. 1000</i>
<i>Pal. 1</i>		100	... 50	... 30	... 170
<i>Pal. 2</i>			... 6	... 110	... 15
...
<i>Pal. i</i>		 25	... 43
...
<i>Pal. j</i>		 50
...
<i>Pal. 1000</i>		

(2)

Como observamos en la ec. (2), C_{ij} es igual a 25, lo que significa que las palabras «i» y «j» aparecen juntas en un total de 25 documentos.

En teoría, a partir de esta matriz de adyacencia podríamos reconstruir completamente la red ciencimétrica que genera el campo científico en estudio; pero en la práctica no es conveniente, ya que los valores de las co-ocurrencias, tal cual, dependen del tamaño de la muestra. Bajo estas circunstancias, los estudios de comparación de redes descritas por diferente número de documentos, serían incorrectos. Es por tanto conveniente recurrir a la normalización de los valores de las co-ocurrencias.

3.2. Medida de los enlaces entre palabras clave: el índice de equivalencia

En la primera versión de Leximappe, la medida normalizada de los enlaces se llevaba a cabo usando los índices de Jaccard y de inclusión [Rip et al., 1984], pero posteriormente se define un nuevo índice más conveniente denominado índice de asociación o de equivalencia [Michelet, 1988]:

$$E_{ij} = \frac{C_{ij}^2}{C_i C_j} \quad (3)$$

donde: E_{ij} .- Índice de equivalencia o de asociación entre las palabras i y j .

C_{ij} .- Co-ocurrencia de las palabras i y j .

C_i .- Ocurrencia de la palabra i .

C_j .- Ocurrencia de la palabra j .

La ecuación anterior se puede reescribir de la siguiente forma:

$$E_{ij} = \frac{C_{ij}}{C_i} \times \frac{C_{ij}}{C_j} \quad (4)$$

donde el primer factor es la probabilidad de tener la palabra i cuando se tiene la palabra j , y el segundo es la probabilidad de tener la palabra j cuando se tiene la palabra i . Este índice es una medida de la relación «Y» entre las palabras i y j .

Los valores de E_{ij} oscilan entre 0 y 1. Cuando dos palabras no aparecen nunca juntas, su co-ocurrencia es nula, el índice de equivalencia vale cero. En cambio, cuando dos palabras siempre que aparecen lo hacen juntas en los mismos documentos, el índice de equivalencia es la unidad. Este índice es independiente del tamaño de la muestra.

3.3. Matriz normalizada de asociaciones

Empleando el índice de equivalencia, la matriz de asociaciones queda normalizada. En nuestro ejemplo nos daría:

	<i>Pal. 1</i>	<i>Pal. 2</i>	...	<i>Pal. i</i>	...	<i>Pal. j</i>	... <i>Pal. 1000</i>
<i>Pal. 1</i>		0.006	...	0.036	...	0.001	... 0.096
<i>Pal. 2</i>			...	0.000	...	0.012	... 0.001
...
<i>Pal. i</i>			0.015	... 0.128
...
<i>Pal. j</i>		 0.014
...
<i>Pal. 1000</i>		

(5)

Si comparamos los valores de esta matriz con la de asociaciones no normalizadas, podemos observar que el par 1,2 que poseía una elevada co-ocurrencia tiene sin embargo un índice de equivalencia bastante menor que el par j,1000 que era de baja co-ocurrencia. Se comprueba, por tanto, que si dos palabras aparecen juntas muchas veces pero proporcionalmente son aún mayores sus ocurrencias por separado, el índice de equivalencia será bajo y Leximappe considerará la unión poco fuerte. En cambio, dos palabras poco frecuentes pero siempre que aparecen lo hacen en los mismos documentos, tendrán un índice de equivalencia muy elevado y por tanto su asociación será muy fuerte. Supongamos, también, una palabra que aparece en muchísimos documentos y que no tiene «predilección» por aparecer conjuntamente con alguna otra en particular sino que se reparte homogéneamente con todas; en este caso, nunca llegará a formar asociaciones consistentes y Leximappe la considerará demasiado genérica y poco significativa.

En definitiva, mediante el uso del índice de equivalencia, Leximappe es capaz de discernir qué palabras y qué asociaciones son realmente relevantes en la construcción de la red ciencimétrica y eliminar aquellas que por su baja co-ocurrencia relativa o su elevada generalidad no lo son.

3.4. Construcción de agrupaciones, temas o subredes: actores temáticos

La matriz de asociaciones normalizada es la matriz de adyacencia del grafo que representa la red. Cada vértice de este grafo es un descriptor y cada índice de equivalencia entre cada dos descriptores es la ponderación de los arcos que unen estas parejas de vértices. En principio sería reconstruible directamente la red imponiendo un umbral mínimo o bien realizar una representación gráfica en dos o tres dimensiones usando un análisis MDS o bien una estructura jerarquizada del tipo dendrograma. Se ha comprobado que posibilidades como éstas no son óptimas para nuestros propósitos [Courtial, 1986], siendo necesario establecer un algoritmo o algoritmos que sean capaces de:

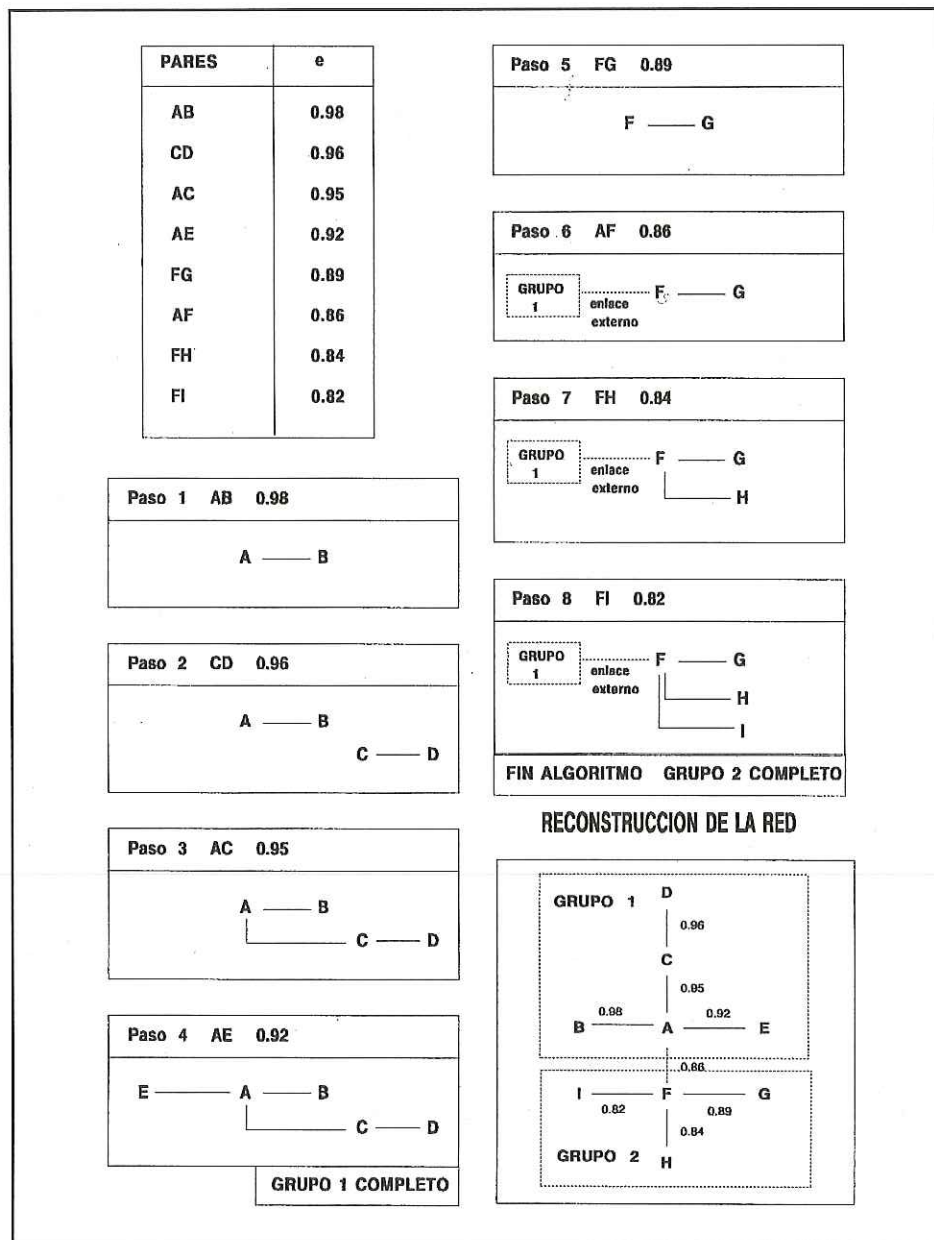
- Extraer de la red ciencimétrica (excesivamente extensa por el elevado número de vértices y enlaces) aquellas agrupaciones o subredes significativas. Esta subredes representarían los temas de investigación y definirían los actores que forman la red global.
- Ofrecer una estabilidad suficiente de los esqueletos (actores-red) frente a factores negativos como errores en la indización y tamaño de la muestra.
- Poder controlar perfectamente las dimensiones de las subredes que definen los actores (número de palabras y umbral de enlace de los temas).
- Capacidad de calcular parámetros que cuantifiquen los actores y los definan como objetos o sujetos y poder seguir su evolución temporal o dinámica.
- En definitiva, que conceptos de la teoría actor-red como las traducciones puedan ser accesibles a un simple cálculo con un microordenador.

Leximappe posee dos algoritmos, ejecutables por sendos módulos independientes: programa COMPS.EXE y programa STAR.EXE. Sus denominaciones son las siguientes: Algoritmo de clasificación por enlace simple (COMPS.EXE) y algoritmo de agrupación por centros simples (STAR.EXE) [Courtial, 1.990] y [Whittaker, 1988].

- a) *Algoritmo de clasificación por enlace simple.* Los elementos de la matriz de asociaciones son ordenados en una lista decreciente según su índice de equivalencia. Esta lista está formada tan sólo por aquellas palabras que tengan una ocurrencia mínima y pares de asociaciones también con una co-ocurrencia mínima preestablecidas. El programa recorre la lista desde el principio y va construyendo dobletes, tripletes, etc. de palabras asociadas de forma que suministra un grafo conexo que no exceda de un valor máximo de palabras preestablecido (por ejemplo 10 ó 15). Cada vez que se obtiene un grafo, elimina las palabras de éste de la lista y comienza el proceso de construcción de nuevos grafos hasta agotar el total de palabras disponibles. La figura 1 es un ejemplo extraído y reelaborado de Courtial (1990) que explica los pasos que se siguen usando este algoritmo para la formación de agrupaciones con un máximo de 5 palabras y la reconstrucción posterior de la red.
- b) *Algoritmo de agrupación sobre centros simples.* Este algoritmo también ordena los pares de asociaciones por orden decreciente de índice de equivalencia y sólo pueden formar parte de esta lista las palabras con una ocurrencia mínima y los pares con una co-ocurrencia mínima establecidas previamente. El ordenador inicializa un contador para cada descriptor y comienza a recorrer la lista desde el principio incrementando el contador de las palabras que van apareciendo. Cuando el contador de una palabra alcanza un valor igual al número de palabras máximo estipulado para los temas menos uno, el algoritmo toma esta palabra como centro de una agrupación. El conjunto resultante estará formado por las uniones de esta palabra central y todas aquellas que se han asociado con ella. El resultado es una estructura en forma de estrella. Las palabras que han aparecido se eliminan de la lista y se comienza de nuevo el proceso para generar más agrupaciones. Si después de recorrer toda la lista ningún contador llega al valor máximo preestablecido, éste se disminuye en tantas unidades como sea necesario para formar una nueva agrupación. El proceso finaliza cuando el valor máximo del contador disminuya hasta un valor mínimo preestablecido o se terminen todas las palabras de la lista ordenada de pares.

Este algoritmo tiene la ventaja, frente al anterior en que nos asegura que cualquier subred obtenida contiene al menos una palabra unida a todas las demás. Esta palabra principal nos va a facilitar la identificación del tema de investigación.


FIGURA 1
ALGORITMO DE CLASIFICACIÓN POR ENLACE SIMPLE



La figura 2 explica el proceso de agrupación sobre centros simples para el ejemplo de la figura 1. Se observa que el resultado es diferente ya que resulta un solo grupo con una estructura en estrella que será identificado mediante la palabra A.

FIGURA 2
ALGORITMO DE AGRUPACIÓN SOBRE CENTROS SIMPLES

PAR	θ	A	B	C	D	E	F	G	H	I
AB	0.98	1	1	0	0	0	0	0	0	0
CD	0.96	1	1	1	1	0	0	0	0	0
AC	0.95	2	1	2	1	0	0	0	0	0
AE	0.92	3	1	2	1	1	0	0	0	0
FG	0.89	3	1	2	1	1	1	1	0	0
AF	0.86	4	1	2	1	1	2	1	0	0



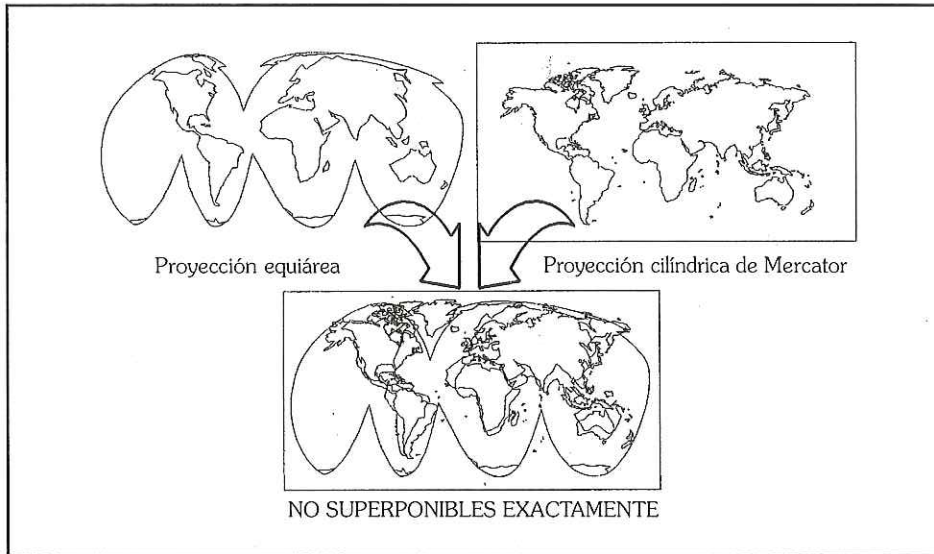
 Palabra principal

			C		
				0.95	
B	—	0.96	A	—	0.92
				0.86	E
			F		

Ambos métodos pueden dar resultados algo diferentes a la hora de definir las subredes, pero al reconstruir la red global, el resultado es siempre el mismo.

La propia esencia de las redes ciencimétricas y sociocognitivas es el de la presencia de fronteras difusas, por lo que no es de extrañar que no sea posible definir las exactamente. Según el algoritmo utilizado, trazaremos más hacia un lado o hacia otro de la frontera difusa la línea divisoria que nos servirá de referencia, pero debe entenderse que esta línea es sencillamente una guía para adentrarnos de forma simplificada en el estudio de las redes que de por sí son extremadamente complejas. Este fenómeno se asemeja al que aparece cuando un geógrafo intenta representar el globo terrestre en un papel. Para ello puede utilizar diversos tipos de proyecciones que no son superponibles exactamente unas sobre otras, presentándose un corrimiento de fronteras y costas. En la figura 3 hemos superpuesto dos mapas de la tierra y es imposible hacerlos coincidir. ¿Cuál de ellos es incorrecto? ¿Lo son los dos? Ninguno de ellos es incorrecto pero tampoco es perfecto. En cualquier caso ambos serán de gran utilidad para un navegante que quiera, por ejemplo, dar la vuelta al mundo.

FIGURA 3
INFLUENCIA DE LA PROYECCIÓN EN LOS MAPAS



4. LOS INDICADORES DE RELACIÓN

4.1. Centralidad y densidad

Una vez identificados los actores o temas, definidos por sus descriptores y por los enlaces que los unen, es conveniente poder establecer parámetros numéricos que de alguna forma nos hagan referencia a sus estructuras internas y a su relación con la globalidad de la red. Se definen los índices siguientes:

- a) Densidad.** La densidad o índice de cohesión interna es la intensidad de las asociaciones internas de un tema y representa el grado de desarrollo que posee. Se calcula como el cociente entre la suma de los índices de equivalencia internos y el número de palabras que definen el tema multiplicado por 100.

$$d=100 \times \frac{\sum_{i=1}^L e_i}{P} \quad (6)$$

donde: e_i .- Índice de equivalencia del enlace interno i

L .- Número de enlaces internos del tema.

P .- Número de palabras del tema.

Por evitar números decimales la densidad suele multiplicarse por 100. Densidades elevadas corresponden a temas altamente desarrollados, muy especializados y repetitivos. Estos temas son actores objeto u objeto-sujeto.

Si ordenamos un conjunto de temas por orden creciente de densidad, el rango de cada tema es lo que se denomina *rango densidad*. Cuando se normaliza, dividiendo entre el número total de temas de la red, presenta valores entre 0 y 1. Se utiliza en la construcción del diagrama estratégico como sinónimo de densidad y es indispensable para hacer estudios comparativos con otras redes y en estudios dinámicos.

$$r_d = \frac{\text{rango}}{N} \quad (7)$$

donde: rango.- Rango del tema según su densidad.

N.- Número de temas de la red.

b) *Centralidad*. La centralidad o índice de cohesión externa es la suma de los índices de equivalencia de todos los enlaces externos que posee un tema. Usualmente el valor de la centralidad se multiplica por 10.

$$c = 10 \times \sum_{j=1}^T e_j \quad (8)$$

donde: e_j .- Índice de equivalencia del enlace externo j

T.- Número total de enlaces externos.

Un tema con elevada centralidad está situado en el centro de la red y se relaciona muy bien con los demás actores. Estos temas son actores sujeto u objeto-sujeto.

Si de forma análoga a la densidad ordenamos un conjunto de temas por orden creciente de centralidad, el rango de cada tema (que puede ser también normalizado dividiendo entre el número total de temas) es lo que se denomina *rango centralidad*. Se utiliza, junto con el rango densidad en la construcción del diagrama estratégico como sinónimo de centralidad y es imprescindible para hacer estudios comparativos entre redes y en estudios dinámicos.

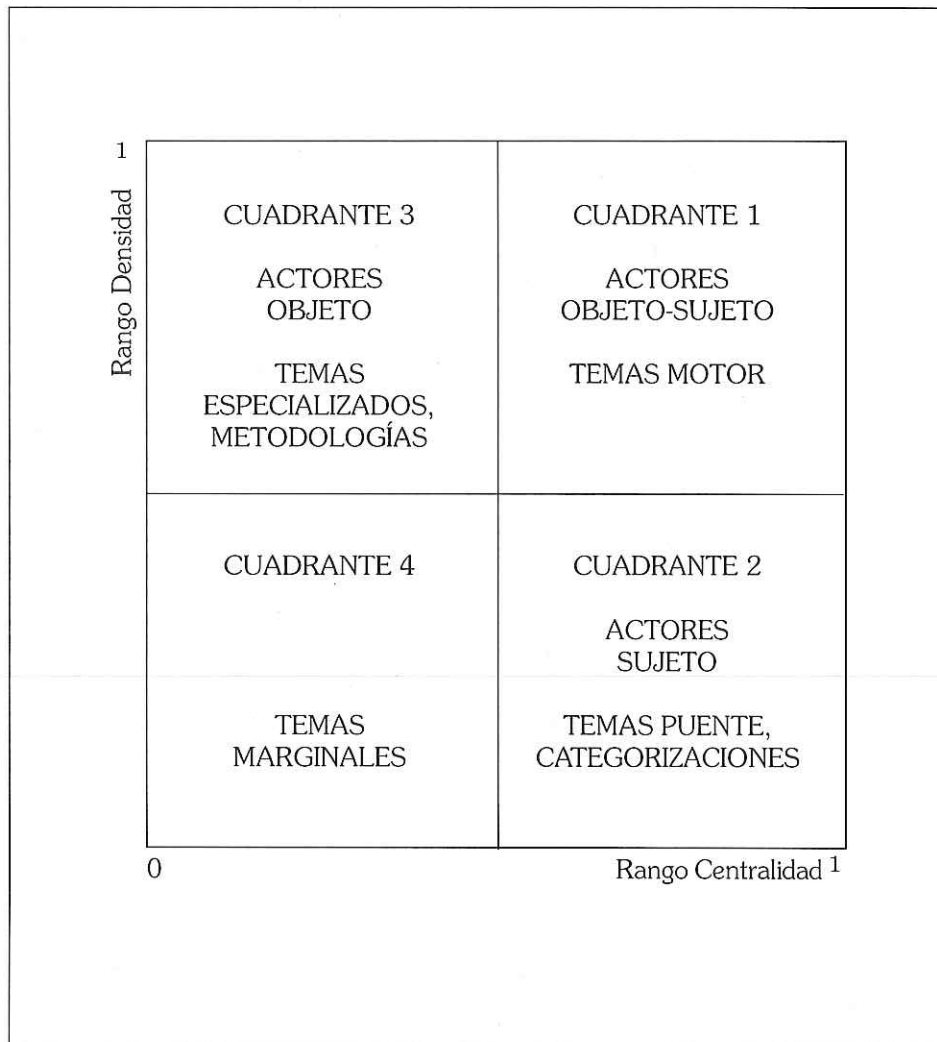
$$r_c = \frac{\text{rango}^1}{N} \quad (9)$$

donde: rango'.- Rango del tema según su centralidad.

4.2. Diagrama estratégico.

Si se representa en un diagrama cartesiano en el eje de abscisas el rango centralidad y en el de ordenadas el rango densidad, se obtiene el llamado diagrama estratégico (Figura 4). El diagrama se divide en cuatro cuadrantes cuyas características son las siguientes:

FIGURA 4
DIAGRAMA ESTRATÉGICO



Primer cuadrante: Situado arriba y a la derecha posee una densidad y centralidad elevadas. Los temas situados en él se caracterizan por estar muy desarrollados y ser centrales. Juegan un papel «motor» dentro del campo científico considerado.

Segundo cuadrante: Baja densidad con alta centralidad. Los temas, bien relacionados pero al tiempo poco desarrollados, pueden considerarse como emergentes o como temas puente.

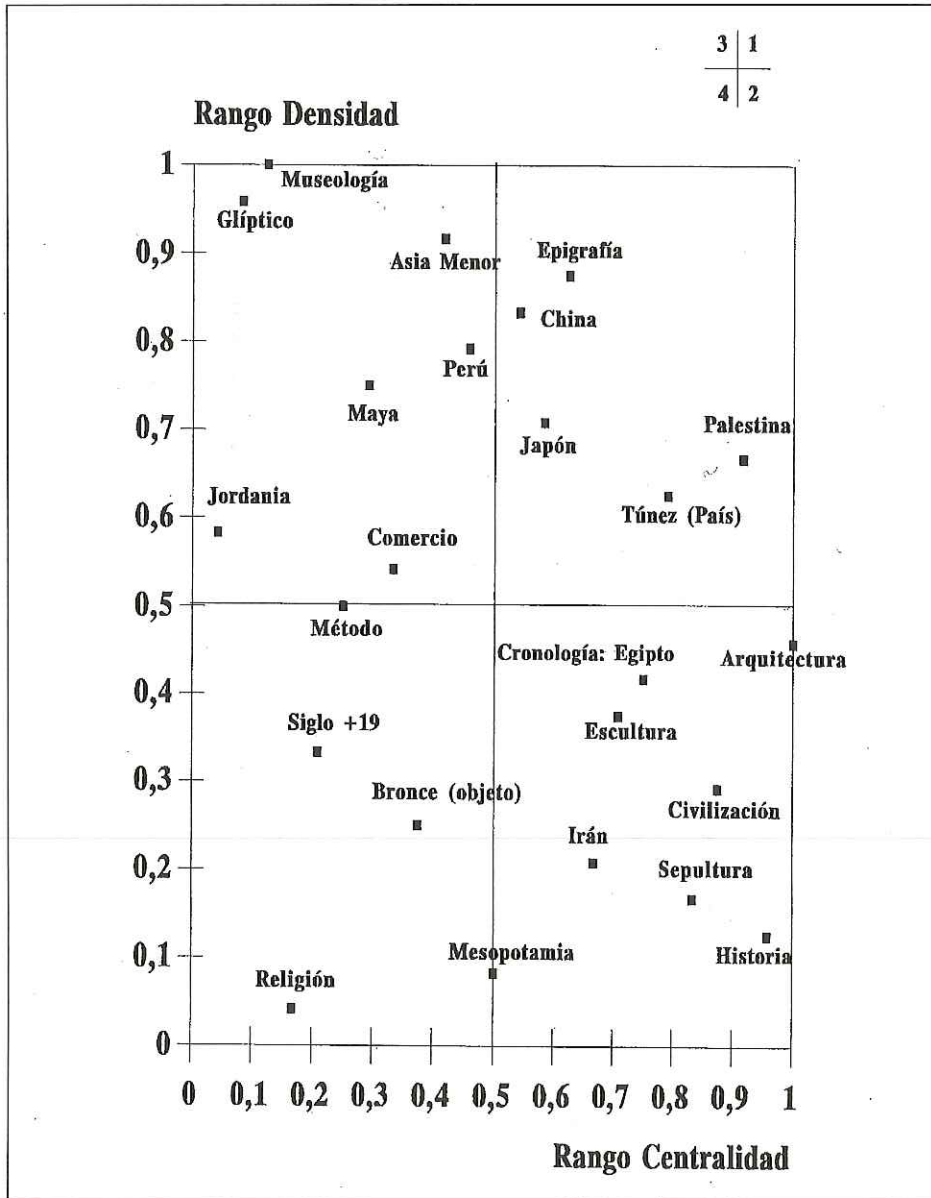
Tercer cuadrante: En él se sitúan los temas muy desarrollados (d elevada) pero poco centrales (c baja). Estos temas pueden considerarse como altamente especializados representativos de una alta actividad, pero aislados en el seno del campo científico en estudio.

Cuarto cuadrante: La centralidad y la densidad son bajas, por lo que los temas aquí situados poseen un carácter débil y netamente marginal. En este cuadrante suelen aparecer por primera vez los temas y también en muchos casos terminan aquí por desaparecer definitivamente.

Hay que aclarar que dos temas próximos en el diagrama estratégico no tienen por qué estar fuertemente ligados desde el punto de vista semántico.

A título de ejemplo, en la figura 5 se muestra el diagrama estratégico de la Arqueología Internacional en 1993. Los temas que aparecen representan el paradigma de investigación de ese año, lo que realmente interesa a los investigadores en ese momento concreto. No aparecen temas como Roma o Grecia debido a que siendo fundamentales para un manual de Arqueología, en el año 1993 no son motivo de una intensa actividad investigadora. En cambio, los temas motores son en este año la Epigrafía, las civilizaciones china y japonesa, la región de Palestina y el norte de África representado por Túnez. Están situados en el centro de la red (elevada centralidad) y están muy bien desarrollados internamente (elevada densidad). Como temas puente aparecen Egipto, Arquitectura, Escultura, Civilización, Irán, Sepultura e Historia. Son temas muy generales, muy bien relacionados y, salvo Egipto e Irán, aplicables prácticamente a cualquier área investigadora. En ellos se establecen las categorías y las clasificaciones en la Arqueología, por ello presentan una elevada centralidad y una baja densidad. En el cuadrante de arriba a la izquierda, los temas de elevada densidad y baja centralidad están representados por el estudio de civilizaciones de menor impacto en el conjunto de la red, pero no por ello dejan de estar muy bien desarrolladas internamente. Se añaden además Glíptica, como instrumento de la epigrafía y Museología como ciencia periférica a la Arqueología pero muy bien desarrollada. Por último comentar que los temas del cuadrante de abajo a la izquierda, en especial Siglo +19, Bronce (objeto) y Religión, pueden considerarse marginales a la Arqueología, lo que no implica que sean centrales y bien desarrollados en otras redes científicas.

FIGURA 5
DIAGRAMA ESTRATÉGICO DE LA ARQUEOLOGÍA INTERNACIONAL
EN 1993



Como se ha observado el diagrama estratégico es una herramienta insustituible para estudiar las posiciones relativas dentro de la red de los actores. Igualmente, en estudios dinámicos nos permite analizar sus traslaciones a lo largo del tiempo.

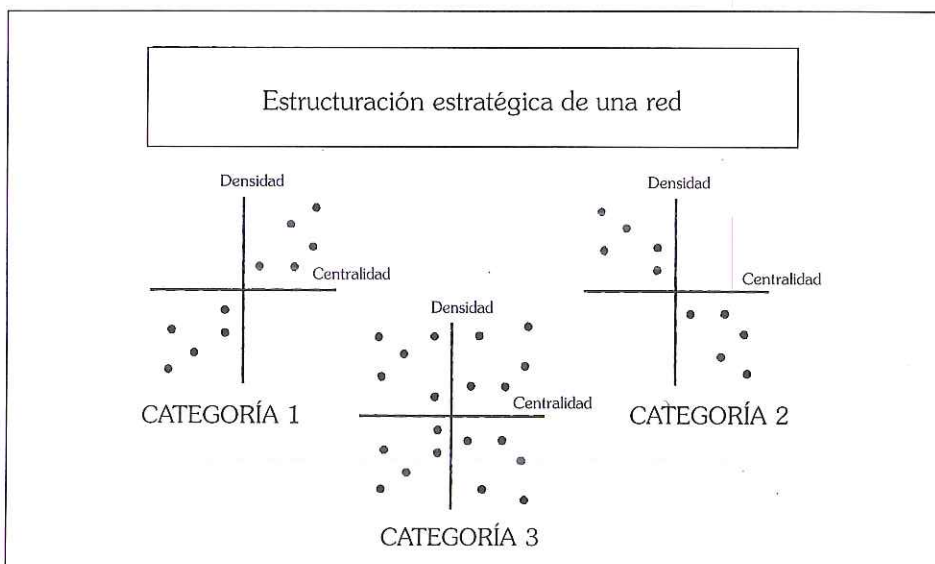
4.3. Reconstrucción de la red

Una vez determinados los temas los podemos asociar desde un punto de vista semántico y estructural de tal forma que es posible reconstruir toda la red del área científica en estudio. Para ello recurrimos al siguiente algoritmo [Rocher et al., 1995]: dos temas están enlazados si existen al menos tres líneas de asociación entre palabras pertenecientes a estos temas y que posean un valor del índice de equivalencia superior al valor más pequeño de las asociaciones internas existentes en ellos.

4.3.1. Estructuración estratégica de la red

El diagrama estratégico nos puede permitir hacer un análisis de la estructuración de la red en función del número relativo de temas que haya en cada cuadrante. Según se distribuyan los temas, pueden considerarse tres tipos de organizaciones [Callon et al., 1995], (Figura 6):

FIGURA 6
CATEGORÍAS EN QUE SE ESTRUCTURA UNA RED



Categoría 1: Los temas se distribuyen alrededor de la bisectriz que une los cuadrantes 1 y 4. Esto indica que la red se organiza en torno a un núcleo de temas bien relacionados y desarrollados y que están en contacto con un conjunto de temas poco desarrollados y periféricos.

Categoría 2: La distribución de temas se realiza en torno a la otra bisectriz formada entre los cuadrantes 2 y 3. Indica que la red está en vías de estructuración o en vías de desintegración ya que apenas hay temas motor. La mayor parte de ellos se distribuyen entre temas especializados y temas puente.

Categoría 3: La distribución es muy homogénea, ya que los temas están repartidos entre los cuatro cuadrantes. La red está entonces muy bien estructurada, es compleja y muy rica, y presenta todos los tipos de temas: los hay centrales, los hay periféricos, unos están bien desarrollados y otros presentan una baja cohesión interna. Una estructuración de este tipo es indicativa de una buena dinámica del campo en estudio.

Callon et al. [1995] recomiendan aplicar índices que nos cuantifiquen en qué grado los temas se acumulan en los diversos cuadrantes. De esta forma, se podría dar una medida comparable de la evolución de la estructuración de una red.

CONCLUSIONES

A lo largo de este trabajo hemos podido comprobar cómo el método de las palabras asociadas es capaz de reconstruir la estructura de una red científica. Con los índices de centralidad y densidad es posible tipificar cada tema de investigación y situarlo respecto de los demás mediante el uso del diagrama estratégico. Además, puede identificar aquéllos que en cada momento son motores en el campo científico considerado, los emergentes y que en un futuro podrán alcanzar un desarrollo notable, los cohesionados pero más periféricos en la red y finalmente los marginales. En definitiva, el método de las palabras asociadas representa una herramienta altamente eficaz para el análisis estructural de cualquier campo científico y un primer acercamiento para su correcta documentación.

BIBLIOGRAFÍA

CALLON et al., 1991.

CALLON, M.; COURTIAL, J.P., y LAVILLE, F. Co-word analysis as a tool for describing the network of interactions between basic and technological research: the case of polymer chemistry. *Scientometrics*, 1991, vol. 22, nº 1, p. 155-205.

CALLON et al., 1995.

CALLON, M.; COURTIAL, J.P., y PENAN, H. *Cienciometría. La medición de la actividad científica: De la bibliometría a la vigilancia tecnológica*. Gijón: Trea, 1995. ISBN 84-87733-94-8.

CHARTRON, 1989.

CHARTRON, G. Lexicon management tools for textual database: the Lexinet system. *Journal of Information Science*, 1989, vol. 15, p. 339-344.

COURTIAL, 1986.

COURTIAL, J.P. Technical issues and developments in methodology. *En*: Callon, M.; Law, J., y Rip, A., eds. Mapping the dynamics of science and technology: Sociology of science in the real world. London: MacMillan Press LTD, 1986.

COURTIAL, 1990.

COURTIAL, J.P. Introduction à la scientométrie: de la bibliométrie à la veille technologique. Paris: Anthropos, 1990. ISBN 2-7178-1935-5.

CORTIAL et al., 1993

COURTIAL, J.P., CALLON, M., y SIGOGNEAU, A. The use of patent titles for identifying the topics of invention and forecasting trends. *Scientometrics*, 1993, vol. 26, nº 2, p. 231-242.

CSI, 1992.

CSI. Centre de Sociologie de L'Innovation. Leximappe-Doc. École Nationale Supérieure des Mines, Paris, France [1992].

MICHELET, 1988.

MICHELET, B. L'analyse des associations [Thèse de doctorat]. Paris, 1988.

RIP, et al., 1984

RIP, A., y COURTIAL, J.P. Co-word maps of biotechnology: an example of cognitive scientometrics. *Scientometrics*, 1984, vol. 6, nº 6, p. 381-400.

ROCHER, et al., 1995

ROCHER, Y.A., y COURTIAL, J.P. (Centre de Sociologie Ecole de Mines de Paris). L'analyse d'un collectif a travers l'exemple d'un domaine de recherche. Le modele de la traduction généralisée dans les interactions sociales, Journées CSI-95. Ecole de Mines de Paris, 11-13, diciembre, 1995.

RUIZ-BAÑOS, 1997

RUIZ BAÑOS, R. Ciencimetría de redes. Análisis de la investigación internacional sobre Arqueología mediante el Método de las Palabras Asociadas (1980-1993) [Tesis Doctoral]. Granada, 1997.

RUIZ-BAÑOS et al., 1997

RUIZ-BAÑOS, R. Y BAILÓN-MORENO, R. Métodos para medir experimentalmente el envejecimiento de la literatura científica. *Boletín de la Asociación Andaluza de Bibliotecarios*, 1997, nº 46, p. 57-75.

WHITTAKER, 1988

WHITTAKER, J. Co-word analysis: The Keele programs. Reino Unido: Universidad de Keele, 1988. Disponible en: Centre de Sociologie de l'Innovation. Ecole National Supérieure des Mines. París.